基于用户间信任关系改进的协同过滤推荐方法*

薛福亮　刘君玲
(天津财经大学商学院　天津 300222)

摘要:【目的】利用用户间信任关系改进协同过滤推荐中用户相似性计算精度, 即在目标用户没有相似用户的前提下, 从其信任用户中选择信任值高的作为相似用户, 进而提高相似用户聚类效果, 提高推荐质量, 并有效缓解协同过滤推荐稀疏性和冷启动问题。【方法】筛选信任用户作为相似用户; 根据选择的信任用户和目标用户形成一个项目的评分集, 并对目标用户未评价过的项目进行评分估算(根据信任用户评分进行简单的评分计算); 将用户间的信任关系依据方差大小进行量化, 形成一个调节因子。本文的创新点就在于调节因子的计算, 并将调节因子纳入用户相似性计算, 形成相似性用户聚类簇, 在此基础上在相似用户之间进行交叉推荐。【结果】通过平均绝对误差指标进行实验评价, 结果表明基于信任关系的协同过滤推荐方法相比传统协同过滤, 在推荐精度上更加准确, 并同时有效缓解了冷启动和稀疏性问题。【局限】本文提出的方法仅在具有信任关系的一个算例上进行实验测试, 需在其他数据集和真实应用场景下进一步检验。【结论】用户间信任关系蕴涵非常有价值的信息, 对用户信任关系进行量化, 并纳入用户相似性计算, 在此基础上实施协同过滤推荐, 对缓解冷启动与稀疏性问题具有较好的理论和实践意义。

关键词: 电子商务推荐　用户信任　协同过滤　冷启动　稀疏性
分类号: TP301.6

1 引　言

推荐系统(Recommender System, RS)是指根据用户的兴趣偏好, 推荐用户感兴趣的对象, 也称个性化推荐系统。

提供高质量的建议对于电子商务系统帮助用户进行选择是很重要的。协同过滤是一种被广泛接受的技术方法, 基于相似性, 生成用户的喜好, 然而, 它具有一些固有的问题, 如数据稀疏问题和冷启动问题。为了解决这些问题, 笔者提出一种新的方法, 即将信任信息合并到相似用户的筛选中, 具体来说, 根据目标用户的信任邻居的评分或偏好推断目标用户的评分和偏好[1-9]。此外, 合并评级的质量是由一个调节因子判定的, 这也是本文的创新点, 对推荐技术的后续发展有非常大的积极作用。

2 研究背景

随着电子商务的日益成熟, 推荐系统的发展也越来越丰富, 适应不同形式、不同行业、不同场景下的推荐系统特点更加鲜明。余力等[10]根据推荐技术分类, 将推荐系统主要分为 6 种: 协同过滤推荐、基于内容推荐、基于人口统计信息推荐、基于效用推荐、基于知识推荐和基于规则推荐, 并对以上 6 种推荐技术进行了评析。

刘建国等[11]则在此基础上提出混合推荐算法, 在协同过滤系统中加入基于内容的算法, 即利用用户的配置

通讯作者: 薛福亮, ORCID: 0000-0002-9344-3489, E-mail: fuliangxue@163.com。

*本文系教育部人文社会科学一般项目“电子商务环境下顾客购物偏好推荐及企业利润挖掘”(项目编号: 13YJC630195)的研究成果之一。

文件进行传统的协同过滤计算。用户的相似度通过基于内容的配置文件计算而得，而非通过简单的用户的评分，这种方法可以克服协同过滤系统中的稀疏性问题。

邓爱林等[12]针对用户评分数据极端稀疏情况下传统相似性度量方法的不足，提出一种基于项目评分预测的协同过滤推荐算法，根据项目之间的相似性初步预测用户对未评分项目的评分，在此基础上，采用一种新颖的相似性度量方法计算目标用户的最近邻居。该算法有效地解决了用户评分数据极端稀疏情况下传统相似性度量方法存在的问题，显著地提高推荐系统的推荐质量。

结合信任的推荐系统可以有效地缓解传统协同过滤算法中存在的数据稀疏问题，并能给每个用户提供可信且准确的推荐。龙宇等[13]在此基础上针对不同用户采用不同推荐模式查找推荐群体，以做出更具个性化的推荐。研究了微观层次上的节点特性，引入兴趣的概念，验证被推荐者的节点特性对于推荐结果的影响。

目前基于信任的推荐算法都是单一的信任模型。邹本友等[14]提出一种基于主题的张量分解的用户信任推荐算法，用来挖掘用户在进行选择时对不同朋友的信任程度。结果表明基于主题的用户信任推荐算法比现有算法具有更好的准确性，并且增量更新的推荐算法可以大幅度提高推荐算法的准确度。

推荐算法的不断创新是为了解决协同过滤推荐技术中两个比较突出的问题，即稀疏问题和冷启动问题。稀疏问题是指有的用户参与的互动活动较少，参与评价的项目很少，这就导致该用户的数据信息很少，并且该用户的相似用户也会很少，所以在对该用户进行协同推荐时准确度就会降低。冷启动问题指新用户还没有参与过任何一个项目的评分，或者新晋的项目还没有被任何一个用户评价过。因为这两个问题的存在，使得协同过滤推荐的结果准确率有所降低。Guo 等[15]提出将一些其他信息添加到协同过滤技术中，这种信息包括亲密关系、会员身份和社会信任，以更好地选择目标用户的相似用户或信任用户，提高推荐的准确度，其认为信任比前两个因素要更有可信度和说服力。该方法可以同时有效缓解稀疏问题和冷启动问题。Ma 等[16]提出一种基于信任和不信任聚类的协同过滤推荐方法。基于 SVD 符号的聚类算法处理信任和不信任关系矩阵，以发现信任社区；并提出一种稀疏的等级补码算法来产生密集的用户评级分布，可以在很大程度上缓解稀疏和冷启动问题。Jia 等[17]提出一种基于多维信任模型的强大的协同过滤推荐算法。从三个方面衡量用户评级的可信度：项目推荐的可靠性、评级相似性和用户的可信度；结合信任计算模型和传统协同过滤方法，选择可靠的邻居集，为目标用户生成推荐；使用 MovieLens 数据集从推荐精度和准确性两方面验证新算法的性能。Xu 等[18]提出一种基于分类和用户信任的改进算法。根据不同项目的评分，对于每个类别，它评估每个用户对该类别的可信程度，并加权计算用户的评分。最后，该算法探讨用户之间的相似性，找到最近邻居，并在每个类别内提出建议。改进的算法优于传统的协同过滤算法，提高了推荐的准确性。Du 等[19]引入社会学领域的信任关系计算，代替传统的相似度计算方法，将信任度集成为最近邻选择。信任网络由不同的路径长度的扩展构成，用户的信任值可以通过信任传递规则获得。

本文主要围绕在协同过滤的方法中添加信任信息以缓解传统方法中数据稀疏问题和冷启动问题，其创新点是在传统的相似值计算过程中添加了一个减少误差的调节因子η，增强了相似值的准确度。

信任信息分为显性信任和隐性信任，本文只对显性信息进行研究。将信任信息“合并”到推荐系统中，以信任为基础确定目标用户的信任用户，如果该目标用户没有相似用户，可以通过确定其信任用户预测目标用户的评分，从信任用户中确定部分相似度较高的用户作为相似用户，从而缓解了推荐系统的稀疏性和冷启动问题。本文提出的方法有效地应用信任用户对目标用户的评分信息进行补值，然后根据补值后的评分数据进一步完成用户相似性计算，确定最相似用户，并最终确定该项目的预测评分。

3　研究框架与方法

本文提出的核心思想如图 1 所示，分为三个阶段：阶段一主要是根据用户信任关系数据库确定信任网络，在目标用户找不到相似用户的情况下，利用“信任可以在信任网络中传播、并且传播距离越远，信任值越小”的特点计算每个信任用户的信任价值，例如，$A \xrightarrow{信任} B, B \xrightarrow{信任} C$，那么 $A \xrightarrow{信任} B \xrightarrow{信任} C$，用 d 表示目标用户到信任用户之间的传递距离，每经

过一次传递 d 值将会加 1。则上述例子中 d_{AB}=1, d_{BC}=1, d_{AC}=2。d 值对信任值是有影响的, d 越大, 信任值越小, 预测的精确度越低。上例中 A 对 B 的信任值大于 A 对 C 的信任值, 考虑到这点, 本文在对信任用户进行筛选时, 一般情况下距离 d 设定为不大于 3 的值。需要说明的是, 这里也有笔者的一个小创新点, 距离为 3 是普遍的一个规定, 如果对于某个项目进行预评分时, 发现距离在 3 以内的信任用户的数量不能完成阶段三的方差计算时(比如当距离在 3 以内的评价过目标项目的信任用户只有一个时, 此时不能用来计算信任用户的评分方差, 应扩大距离范围, 寻找更多评价过目标项目的信任用户完成计算), 因此会考虑 d>3 的信任用户进行预评分。在确定了信任用户的基础上, 阶段二找到所有信任用户以及目标用户评价过的项目, 并生成一个项目集, 对于每个目标用户未进行评价的项目, 根据信任邻居的评分, 利用一个带有权重的均值计算作为目标用户的评分, 权重和阶段一中信任网络中的距离有关。这样直到所有的未评分项目都有一个估算值, 就形成一个新的、完整的评分数据集合。阶段三将针对每个项目的估算值进行方差的计算, 通过方差计算出一个调节因子, 用来对后续的相似值进行调节, 然后根据公式计算出相似值, 根据相似值确定相似用户, 根据相似用户进行评分估算, 进而实现推荐。

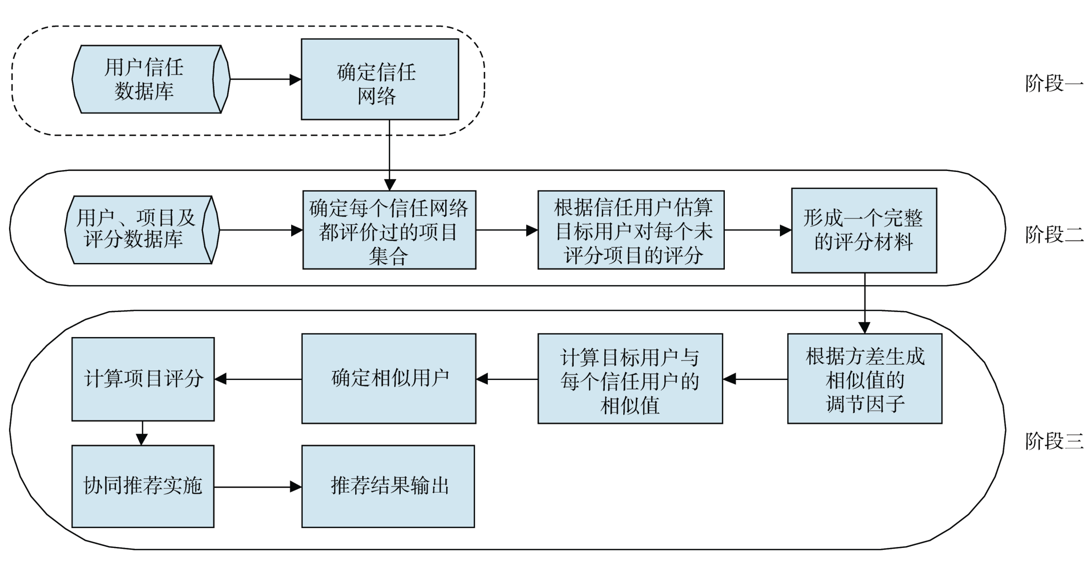


图 1　引入用户信任关系后的协同推荐核心框架

4　研究过程

本文预先在模型中约定变量定义如下: 定义全部的用户集合为 U, 定义全部的项目集合为 I, 所有的评分集合为 R 。u,v 代表用户, i,j 代表项目, r 代表评分, r 的取值范围为[1, 5]的整数。这样, 用户 u 对项目 i 的评分, 可以用一个三维的数组表示为 $(u,i,r_{u,i})$, 这样求目标用户对某个项目的评分可以表示为 $(u,j,?)$ 。另外, 本文定义信任用户的集合为 TN, 所以用户 u 的信任用户集合为 TN_u, 如果目标用户 u 信任用户 v, 那么 $v\in TN_u$ 。设目标用户 u 对用户 v 的信任值为 t, $t\in\{0,1\}$, 代表用户 u 对用户 v 的信任程度, t 值越大说明信任程度越大, 评分越接近。这里说明一下, 用户 u 本身也是自己的信任用户, $u\in TN_u$, $t_{u,u}=1$ 。除此之外, 用户 u 评分过的项目集合用 I_u 表示, $I_u=\{i\,|\,r_{u,i}\in R, i\in I\}$, 对项目 i 进行过评分的用户集合用 U_i 表示, $U_i=\{u\,|\,r_{u,i}\in R, u\in U\}$ 。

推荐问题可以描述为: 给定一组用户评分 $(u,i,r_{u,i})$ 和一组用户的信任 $(u,v,t_{u,v})$, 预测一个最佳预测 $(u,j,\hat{u}_j)$ 为一个目标用户对目标项目 j 的评分。通常通过预测精度评价推荐质量, 即预测结果和真正喜好的接近程度。

4.1　确定目标用户信任用户集合

通过筛选确定目标用户的信任用户。如果一个用户评价积极, 那么确定该用户的相似用户相对容易,

但是，如果用户评价数据过少，这样他的相似用户往往很难确定。因为不活跃用户的相似用户数量大多数不能满足预测评分要求的数量。利用相似用户估算出的结果肯定要比利用信任用户计算的结果更准确。本文研究的是后者，即当目标用户的相似用户的数量过少时，考虑其信任用户进行推荐。

一般定义参与评分项目数少于5的为不积极用户，对不积极用户将寻找其信任用户。虽然不积极用户的信任信息的可用性是相对有限的，但信任可以随着信任的网络传播，即如果 A 信任 B, B 信任 C，那么 A 信任 C，即一个不活跃的用户只要有一个信任用户，那么就可以找到满足数量条件的更多的信任用户(间接信任用户)。从第 3 节可以看出，d 值对信任值 t 值是有影响的，d 越大，t 越小，预测的精确度越低。在本文中，关于 d 对 t 的影响，笔者定义两者是反比关系。即：

$$t_{u,v}=\frac{1}{d_{u,v}} \tag{1}$$

所以信任用户的确定过程如下：

$$TU_U=\{v\mid d_{u,v}\leqslant d, v\in U\} \tag{2}$$

需要说明：关于公式(1)和公式(2)，根据六度分隔理论，在信任网络中任意两个用户可以用小于 6 的步骤连接。所以一般在预测之前都要对 d 设定一个阈值。本文相关计算中，$d\leqslant 3$。

4.2 形成完整的评分数据集

在确定了信任用户后，可以确定一些评分项目作为模型计算的候选项目。

(1) 定义项目集 $\tilde{I}_u$

定义目标用户 u 的所有信任用户(包含目标用户本身)评价过的项目为 $\tilde{I}_u$。

(2) 定义评分集 $\tilde{r}_{u,j}$

定义根据所有的信任用户来确定目标用户 u 对项目 j 的评分为 $\tilde{r}_{u,j}$。

$$\tilde{r}_{u,j}=\frac{\sum_{v\in TN_u} t_{u,v}\cdot r_{v,j}}{\sum_{v\in TN_u}|t_{u,v}|} \tag{3}$$

(3) 皮尔森相关系数计算

利用皮尔森相关系数计算用户相似性的公式如下。

$$s_{u,v}=\frac{\sum_{i\in I_{u,v}}(r_{u,i}-\bar{r}_u)(r_{v,i}-\bar{r}_v)}{\sqrt{\sum_{i\in I_{u,v}}(r_{u,i}-\bar{r}_u)^2}\sqrt{\sum_{i\in I_{u,v}}(r_{v,i}-\bar{r}_v)^2}} \tag{4}$$

其中，$s_{u,v}\in[-1,1]$。$s_{u,v}$=1，表明信任用户与目标用户喜好完全一致；$s_{u,v}$>0，表明信任用户 v 和目标用户 u 之间有积极的联系；$s_{u,v}$<0，表明信任用户 v 和目标用户 u 之间有消极的联系，对方法无意义，本文不考虑。

文献[15]在传统的皮尔森相关系数计算中加入因子 $c_{u,i}$ [20-21]，含义为：在确定目标用户的信任用户的前提下，针对某一个待评分的项目，挑选出对该项目评分的信任用户，根据每个信任用户的打分情况将其分为积极用户和消极用户，积极用户和消极用户的划分方法为：在一个项目 i 中，在所有信任用户评分中确定一个中位数 r_{med} (或许是均值)，评分大于 r_{med} 为积极用户，评分小于 r_{med} 为消极用户。然后将积极用户数量和消极用户数量经过公式量化后添加到相似值的计算中，这样可以在计算目标用户评分时，为评分的积极与否的判定中，添加制衡因子，减少概率误差。该方法的不足在于，当参与评分信任用户全部为积极用户(或者全部为消极用户)时，调节因子不会起作用，假如全部为积极用户(消极用户)，而目标用户实际的偏好是消极的(积极的)，此时在一定程度上对计算结果会有误导作用。

(4) 改进皮尔森相关系数

针对以上问题，本文将公式(4)进一步优化：

$$s''_{u,v}=\frac{\sum_{i\in I_{u,v}}\eta_{u,i}(\tilde{r}_{u,i}-\bar{r}_u)(r_{v,i}-\bar{r}_v)}{\sqrt{\sum_{i\in I_{u,v}}\eta^2_{u,i}(\tilde{r}_{u,i}-\bar{r}_u)^2}\sqrt{\sum_{i\in I_{u,v}}(r_{v,i}-\bar{r}_v)^2}} \tag{5}$$

同样，$s''_{u,v}\in[-1,1]$。$s''_{u,v}=1$，表明信任用户与目标用户喜好完全一致；$s''_{u,v}$>0，表明信任用户 v 和目标用户 u 之间有积极的联系；$s''_{u,v}$<0，表明信任用户 v 和目标用户 u 之间有消极的联系，对方法无意义，本文不考虑。

(5) 选定调节因子 $\eta_{u,i}$

$$\eta_{u,i}=1-\frac{\sqrt{e^2_{u,i}}}{N} \tag{6}$$

η 含义如下：确定信任用户，针对待评分项目确定对其进行过评分的信任用户；对所有参与评分的信

任用户的评分进行方差分析; 将其添加到相似值的计算中。该调节因子分别对每个评分项进行方差分析, 方差降低了误差, 也就意味着在进行相似值计算之前, 针对每个项目的评分都做了减少误差的计算, 这样再计算相似值就大大减少了数据来源本身带来的误差, 增强了相似值的准确度。其中 $e^2_{u,i}$ 是指: 对于目标用户 u 未评价的项目 i, 利用信任用户先计算出信任用户的均值作为 u 对项目 i 的估算值, 然后计算出其方差, 作为误差分析的项。N 为信任用户中评价过当前项目的其他用户的数量。

$$e^2_{u,i}=\frac{\sum_{v\in TN_u,\ n-1}^{N}(r_{v,i}-\overline{r}_{u,i})^2}{N-1} \tag{7}$$

由公式(6)和公式(7)可以看出, 方差的存在本身就起到减少误差的作用, 方差越大, 计算出来的调节因子越小。公式(5)在考虑到信任用户的数量对模型影响的同时, 也考虑到信任用户的评分对模型的影响。

4.3 评分估算

(1) 确定最相近的信任用户 MTU_u

根据公式(5)计算所有信任用户和目标用户 u 的相似性, 再从信任用户中选出满足条件的最信任用户 MTU_u。确定 MTU_u 是给 $s'_{u,v}$ 设定一个阈值 θ_s, 在信任用户中选择所有的 $s'_{u,v}$ 大于 θ_s 的用户作为 MTU_u。即:

$$MTU_u=\{v\mid s'_{u,v}>\theta_s, v\in U\}$$

(2) 计算目标用户 u 对未评分项目 j 的评分

结合确定的 MTU_u, 根据公式(5)计算出的相似值, 可以对要评分的项目进行评分计算。评分结果 $\hat{r}_{u,j}$ 的计算方法如下:

$$\hat{r}_{u,j}=\frac{\sum_{v\in MTU_u}s''_{u,v}\cdot r_{v,j}}{\sum_{v\in MTU_u}\left|s''_{u,v}\right|} \tag{8}$$

就此, 得出模型的最终结果。

5 应用算例

应用一个算例, 利用上面介绍的方法循序渐进地对每一步进行运算, 最后生成项目预测评分。需要说明的是, 算例的选取是有科学的方法的, 并不是随手写的几个关系表和评分表, 而笔者的思维并没有专业人士全面周到, 各个方面都要求不偏离实际, 因此本文的算例是笔者在阅读了大量相关文献后, 斟酌借用的经典算例[6]。

在该算例中, 有 9 个用户, 9 个项目, 分别用 u_k 和 i_j 表示, 其中 $k, j\in[1,9]$, 每个用户对每个项目的评分为 $r_{u,j}$, $r_{u,j}\in[1,5]$且为整数, 如表 1 所示。同时每个用户同其他用户可能存在信任关系, 如表 2 所示, 其中竖列表示信任用户, 横行表示被信任用户, 例如, 横行为 u_1, 竖列为 u_2 的值为 1, 可以写为(u_1, u_2, 1), 表示 u_1 信任 u_2。在该算例里要预测的是 u_1 对 i_5 的评分。

从表 1 中可以看出, u_1 只对 i_3 进行了评分, 且分数为 5。从信任网络(表 2)中看出, u_1 的直接信任用户为 u_2 和 u_3, 也就是说 $d(u_1,u_2)=1, d(u_1,u_3)=1$, 本文设在信任网络中, 与信任用户距离 d 为 1 的用户为一级信任用户, d 为 2 的用户为二级信任用户, 例如 $d(u_1,u_4)=d(u_1,u_2)+d(u_2,u_4)=2$, 依次类推, d 为 3 的用户为三级信任用户。而信任是有方向的, u_1 信任 u_2 并不代表 u_2 也信任 u_1, 所以在信任网络中, 信任关系的传播是单向的。

表 1 用户对项目的评分

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9
u_1			5						
u_2	5		4		3			2	
u_3		4		3				1	
u_4	3		5		2				
u_5		4	4		3			3	
u_6		3	3	5	5				
u_7							5		4
u_8			4		2			1	
u_9			4		5			5	

表 2　用户与用户之间的信任关系

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9
u_1		1	1						
u_2			1	1					
u_3		1		1					
u_4					1				
u_5				1		1			
u_6			1	1					
u_7									
u_8									
u_9									

添加信任方法如下:

(1) 通过信任在信任网络中可以传播的性质确定信任用户。根据表 2, 确定信任值可通过公式(1)推断出来, 结果如表 3 所示。作为一名活跃用户, u_1 总是完全信任它自己, 即 $t_{u_1,u_1}=1.0$ (表示百分百信任), 在此基础上, 进行 u_1 与其他信任用户的信任值的计算。u_2, u_3 是目标用户 u_1 的一级信任用户(d 值都为 1), 这里他们的信任价值也为 1。对于用户 u_4, 距离 u_1 的最小距离 d=2, 即最短的传播路径为 $u_1 \to u_2(u_3) \to u_4$, 其他传播路径, 例如, $u_1 \to u_2 \to\ u_3 \to u_4$, $d=3>d=2$, 在这里选择 d 最短的路径计算信任值 t。所以 $t_{u_1,u_4}=1/2=0.5$。最短的距离到 u_5 为: $d(u_1,u_5)=d(u_1,u_4)+d(u_4,u_5)=2+1=3$。到 u_6 的距离以此类推, 虽然该传播距离可以计算, 但是 $d(u_1,u_6)=4$, 前面已经声明, 本文的传播距离设定为 d≤3。所以 u_6 在本文中不属于目标用户 u_1 的信任用户。但是, 距离为 3 是普遍的一个规定, 如果对于某个项目进行预评分时, 发现距离在 3 以内的信任用户的数量不能完成方差计算时(比如当距离在 3 以内的评价过目标项目的信任用户只有一个时, 此时不能计算信任用户的评分方差, 应扩大距离范围, 寻找更多评价过目标项目的信任用户完成计算), 一般会考虑 $d>3$ 的信任用户进行预评分。例如, 本算例中, 对在 η_{u_1,i_4} 的计算中, 因为 TU_1 中只有 u_3 对 i_4 有评分, 所以在计算调节因子的时候借助距离 u_1 为 4 的 u_6 的评分, 结果如表 4 所示。可以得到 u_1 的信任用户集合 $TU_{u_1}=\{u_1,u_2,u_3,u_4,u_5,u_8\}$。

(2) 合并信任用户的评分, 结果如表 4 所示。此过程一直持续到至少所有信任用户所参与的评分项目(即 $\tilde{I}_u$)全被覆盖, 则一个新的评分配置文件形成。合并后的目标用户 u_1 的评分材料比原始数据要完整得多。

表 3　u_1 和信任网络中的信任用户的信任值

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9
d	0	1	1	2	3	4			
$t_{u1,uk}$	1.00	1.00	1.00	0.50	0.33	0.25			

表 4　合并后的 u_1 的评分

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9
$\widetilde{r}_{u_1,i_j}$	4.33	4	5	3	2.73			1.71	
η_{u_1,i_j}	0.29	1.00	1.00	0.29	0.81			0.67	

(3) 利用前面新的评分材料, 分别计算目标用户 u_1 和每个信任用户 u_k 的相似值 s (公式(4))、s' (添加调节因子 $c_{u,i}$ 的相似值)、s'' (公式(5)), 结果如表 5 所示。由结果看出 u_2, u_3, u_4, u_5, u_8 与目标用户 u_1 的相似值大于 0, 且比较高, 因为要估算的是 u_1 对 i_5 的评分, 而 u_3 并没有对项目 i_3 进行评分, 所以 u_3 并不在集合中。最相似用户的集合 $NN_{u1}=\{u_2,u_4,u_5,u_8\}$。根据公式(8)计算最后的评分结果。

表 5　u_1 和信任用户的相似值

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9
S	1.0	0.870	0.992	0.910	0.910	–0.910		1.0	–0.950
S'	1.0	0.660	0.995	0.980	0.840	–0.780		0.990	–0.980
S''	1.0	0.699	0.955	0.985	0.910	–0.828		0.982	–0.968

$$\hat{r}(s)_{u_1,i_5}=\frac{3\times 0.87+2\times 0.91+3\times 0.91+2\times 1}{0.87+0.91+0.91+1}\approx 2.48$$

$$\hat{r}(s'')_{u_1,i_5}=\frac{3\times 0.699+2\times 0.985+3\times 0.910+2\times 0.982}{0.699+0.985+0.910+0.982}\approx 2.45$$

需要说明：前文提到 u_1 的信任用户集合为 $TU_{u_1}=\{u_1,u_2,u_3,u_4,u_5,u_8\}$，所以表 5 中计算的相似值只包含 u_2,u_3,u_4,u_5和u_8 等 5 个信任用户的相似值即可，包含 u_6 和 u_9 的相似值是为了与信任用户形成对比。因此表 5 不包含 u_7 的对应值。

6　结果及讨论

为了验证该方法的有效性，对一个实际数据集进行实验。具体来说，目的是找出与其他同类方法相比，本文方法的性能如何；信任传播对本文方法和其他方法的影响。

6.1　实验设置

在实验中，比较了本文的方法与文献[15]的添加信任的协同过滤方法(CCF)以及传统的基于用户的协同过滤方法(CF)的性能。

传统的协同过滤方法使用皮尔森相关系数(简称PCC)测量计算用户相似度，选择相似度高于阈值的用户[22-31]。并使用他们的评分生成项目预测。在这项工作中，所有方法的阈值设置为 0。

6.2　评估指标

(1) 评估指标 *MAE*

平均绝对误差或 *MAE*，衡量预测接近实际的程度：

$$MAE=\frac{\sum_u\sum_i\left|\hat{r}_{u,i}-r_{u,i}\right|}{N}\qquad(9)$$

其中，N 是测试级别的数量。因此，*MAE* 值越小，预测越接近实际。文献[20]根据均方根误差(*RMAE*)定义测量精度，笔者定义反向 *MAE* 或 *iMAE* 作为通过等级标度范围归一化的预测精度：

$$iMAE=1-\frac{MAE}{r_{\max}-r_{\min}}\qquad(10)$$

其中，$r_{\max}$ 和 $r_{\min}$ 分别是由推荐器系统定义的最大额定值和最小额定值。较高的 *iMAE* 值表示更好的预测精度。

(2) 评估指标 *F*1

同时考虑到评估的准确性和覆盖率，*F* 度量或 *F*1 可以衡量整体绩效。准确性和覆盖率都是预测性能的重要措施。根据文献[20]，*F* 度量计算方法如下：

$$F1=\frac{2\cdot iMAE\cdot RC}{iMAE+RC}\qquad(11)$$

因此，F-measure 反映了准确性和覆盖率之间的平衡。

6.3　结果与分析

通过算例利用以上的两个指标进行度量，预测性能如表 6 所示。

表 6　算例上的预测性能表

指标	CF	CCF	ECF
iMAE	0.9985	0.9986	0.9987
F1	0.7994	0.7995	0.7996

(说明：ECF 是指本文加入调节因子 $\eta_{u,i}$ 的协同过滤方法。)

因为本算例中数据集较小，所以计算出的两个指标值差距并不明显，但是对于单位为 1 的各指标来讲，这样的差距虽然渺小，但还是具有一定的代表性。后期笔者会用实际数据集进行实验，以完善本文方法。在传统的协同过滤算法中，如果目标用户没有相似用户则无法对其进行评分预测，本算例中所有用户的关系都是信任关系，并不是相似关系，所以以上关于 CF 的各个指标的计算的结果都是用本文方法确定信任用户作为相似用户后进行预测的。从表 6 两个指标可以看出，本文方法获得了更好的准确性和覆盖率。传统的 CF 方法和优化后的方法比较，在误差项分析和预测准确性上都是有微小差距的。而和 CCF 方法相比较，虽然在误差分析上差异并不大，但是在整体绩效的衡量中，ECF 方法略微胜于 CCF 方法。

相对于其他方法，本文方法有两个明显的优势。首先，它可以有效地改善数据稀疏性和冷启动问题。

这两个问题的本质是两个用户共同评价过的项目很少，想要准确地计算用户相似性，并因此找到可靠的类似的用户更加困难，或者，两个用户之间根本没有共同评价过的项目，导致无法计算用户相似值。针对上述问题，本文通过应用信任邻居的评分形成一个新的评分材料代表目标用户的偏好。对于目标用户来说，新形成的评分材料比原始的评分材料覆盖了更多项目，因此，在用户相似性方面，更多的相似用户将会被识别，特别对于那些有很少评分或者没有评分的不活跃用户来说，更有意义。前面的例子也说明考虑评分估算标准差的相似值计算比其他的方法更可靠。

结果表明，本文的方法能有效缓解数据稀疏问题和冷启动问题。另外，当一个不活跃用户或者一个新的用户从来没有参与过任何项目的评价时，但是已经指明了信任用户，这时可以将这些信任用户的评分进行合并形成一个新的评分数据集，作为目标用户的评分。

7 结 语

本文提出一种将可信邻居纳入传统协同过滤技术的新方法，旨在解决传统推荐系统存在的数据稀疏和冷启动问题。纳入信任邻居的评分，以补充和表示目标用户的偏好，进而实现下一步评分预测。在此基础上通过一个算例对数据集进行检验，结果表明，在准确性和整体性能方面，本文方法都取得了微小的进步。此外，通过在信任网络中传播信任，可以实现更好的预测性能。

后续工作中，将采用真实网站上的数据集对该方法进一步测试，验证其在真实环境下的可行性，并进一步推断用户行为的隐性信任，增强信任关系在协同过滤推荐中的适用性。这里说明一下对实际数据集的几点要求:

(1) 数据集包含的样本数量要足够大，最好是千或万的数量单位;

(2) 数据集要包含一定数量的不积极用户，这些不积极用户中有一部分是有很少或者没有相似用户的;

(3) 最重要的一点，数据集中的用户之间存在信任关系。

参考文献:

[1] Wang Y, Singh M P. Formal Trust Model for Multiagent System[C]//Proceedings of the 20th International Joint Conference on Artificial Intelligence.2007: 1551-1556.

[2] Lampropoulos A S, Lampropoulos P S, Tsihrintzis G A. A Cascade-Hybrid Music Recommender System for Mobile Services Based on Musical Genre Classification and Personality Diagnosis[J]. Multimedia Tools and Applications, 2012, 59(1): 241-258.

[3] Shambour Q, Lu J. A Trust-semantic Fusion-based Recommendation Approach for E-business Application [J]. Decision Support Systems, 2012, 54(1): 768-780.

[4] Adomavicius G, Tuzhilin A.Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-art and Possible Extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.

[5] Jøsang A, Quattrociocchi W, Karabeg D. Taste and Trust[C]//Proceedings of IFIP International Conference on Trust Management.2011: 312-322.

[6] Chowdhury M, Thomo A, Wadge W W. Trust-based Infinitesimals for Enhanced Collaborative Filtering [C]// Proceedings of the 15th International Conference on Management of Data. 2009.

[7] Danis C, Singer D. A Wiki Instance in the Enterprise: Opportunities, Concerns and Reality[C]//Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work.2008: 495-504.

[8] De Rosa C, Cantrell J, Havens A, et al. Sharing, Privacy and Trust in Our Networked World[R]. A Report to the OCLC Membership, OCLC, 2007.

[9] Golbeck J A.Computing and Applying Trust in Web-Based Social Networks[D]. College Park, MD, USA: University of Maryland at College Park, 2005.

[10] 余力，刘鲁. 电子商务个性化推荐研究[J]. 计算机集成制造系统，2004, 10(10): 1306-1313. (Yu Li, Liu Lu. Research on Personalized Recommendations in E-business [J]. Computer Integrated Manufacturing Systems, 2004, 10(10): 1306-1313.)

[11] 刘建国，周涛，汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学进展，2009, 19(1): 1-15. (Liu Jianguo, Zhou Tao, Wang Binghong. Study on the Research of Personalized Recommendation System [J]. Progress in Natural Science, 2009, 19(1): 1-15.)

[12] 邓爱林，朱扬勇，施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报，2003, 14(9): 1621-1628. (Deng Ailin, Zhu Yangyong, Shi Bole. A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction [J]. Journal of Software, 2003, 14(9): 1621-1628.)

[13] 龙宇, 童向荣. 结合信任的推荐系统的性质[J]. 计算机应用, 2014, 34(1): 222-226, 235. (Long Yu, Tong Xiangrong. Property of Trust-based Recommendation System [J]. Journal of Computer Applications, 2014, 34(1): 222-226, 235.)

[14] 邹本友, 李翠平, 谭力文, 等. 基于用户信任和张量分解的社会网络推荐[J]. 软件学报, 2014, 25(12): 2852-2864. (Zou Benyou, Li Cuiping, Tan Liwen, et al. Social Recommendations Based on User Trust and Tensor Factorization [J]. Journal of Software, 2014, 25(12): 2852-2864.)

[15] Guo G, Zhang J, Thalmann D. Merging Trust in Collaborative Filtering to Alleviate Data Sparsity and Cold Start [J]. Knowledge-Based Systems, 2014, 57: 57-68.

[16] Ma X, Lu H, Gan Z, et al. An Explicit Trust and Distrust Clustering Based Collaborative Filtering Recommendation Approach [J]. Electronic Commerce Research and Applications, 2017, 25: 29-39.

[17] Jia D, Zhang F, Liu S. A Robust Collaborative Filtering Recommendation Algorithm Based on Multidimensional Trust Model[J]. Journal of Software, 2013, 8(1): 11-18.

[18] Xu X L, Xu G L. Improved Collaborative Filtering Recommendation Based on Classification and User Trust[J]. Journal of Electronic Science and Technology, 2016, 14(1): 25-31.

[19] Du Y, Du X, Huang L. Improve the Collaborative Filtering Recommender System Performance by Trust Network Construction[J]. Chinese Journal of Electronics, 2016, 25(3): 418-423.

[20] Jamali M, Ester M. Trustwalker: A Random Walk Model for Combining Trust-based and Item-based Recommendation [C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009: 397-406.

[21] Diaz-Aviles E, Drumond L, Schmidt-Thieme L, et al. Real-time Top-n Recommendation in Social Streams[C]// Proceedings of the 6th ACM Conference on Recommender Systems.2012: 59-66.

[22] 陈宇亮, 沈奎林. 基于读者评论的图书推荐系统研究[J]. 图书情报导刊, 2016, 1(9): 6-9. (Chen Yuliang, Shen Kuilin. Study on Book Recommendation System Based on Reader’s Comments [J]. Journal of Library and Information Science, 2016, 1(9): 6-9.)

[23] 李琦. 基于社交网络好友信任度的个性化推荐系统研究[D]. 哈尔滨: 哈尔滨工业大学, 2014. (Li Qi. Study on Personalized Recommendation System Based on Social Network Friends’ Trust [D]. Harbin: Harbin Institute of Technology, 2014.)

[24] 肖志宇. 基于社交网络和信任模型的推荐系统的研究与实现[D]. 南京: 东南大学, 2015. (Xiao Zhiyu. Research and Implementation of Recommendation System Based on Social Network and Trust Model [D]. Nanjing: Southeast University, 2015.)

[25] 孙国豪. 社交网络中基于信任的推荐系统[D]. 苏州: 苏州大学, 2015. (Sun Guohao. Trust-based Recommendation System in Social Network [D]. Suzhou: Soochow University, 2015.)

[26] 朱岩, 林泽楠. 电子商务中的个性化推荐方法评述[J]. 中国软科学, 2009(2): 183-192. (Zhu Yan, Lin Zenan. A Review of E-Business Recommendation System [J]. Chinese Soft Science, 2009(2): 183-192.)

[27] 李聪, 梁昌勇, 董珂. 基于项目类别相似性的协同过滤推荐算法[J]. 合肥工业大学学报: 自然科学版, 2008, 31(3): 360-363. (Li Cong, Liang Changyong, Dong Ke. A Collaborative Filtering Recommendation Algorithm Based on Item Category Similarity[J]. Journal of Hefei University of Technology: Natural Science Edition, 2008, 31(3): 360-363.)

[28] 李晓昀, 阳小华, 余颖. 基于隐性反馈分析的个性化推荐研究[J]. 计算机工程与设计, 2009, 30(16): 3794-3796, 3825. (Li Xiaoyun, Yang Xiaohua, Yu Ying. Research on Individualized Recommendation Based on Implicit Feedback Analyses [J]. Computer Engineering and Design, 2009, 30(16): 3794-3796, 3825.)

[29] 余力, 刘鲁, 李雪峰. 用户多兴趣下的个性化推荐算法研究[J]. 计算机集成制造系统, 2004, 10(12): 1610-1615. (Yu Li, Liu Lu, Li Xuefeng. Research on Personalized Recommendation Algorithm for Users Multiple-interests[J]. Computer Integrated Manufacturing Systems, 2004, 10(12): 1610-1615.)

[30] 孙小华. 协同过滤系统的稀疏性与冷启动问题研究[D]. 杭州: 浙江大学, 2005. (Sun Xiaohua. Study on Sparsity and Cold Start of Collaborative Filtering System [D]. Hangzhou: Zhejiang University, 2005.)

[31] 许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(2): 350-362. (Xu Hailing, Wu Xiao, Li Xiaodong, et al. Comparison Study of Internet Recommendation System [J]. Journal of Software, 2009, 20(2): 350-362.)

作者贡献声明:

薛福亮: 提出研究思路, 设计研究方案;

刘君玲: 采集、分析、处理数据, 进行实验评价, 论文起草;

薛福亮, 刘君玲: 论文最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: 408925113@qq.com。

[1] 薛福亮, 刘君玲. user's similarity data.txt. 用户的相似值数据.

[2] 薛福亮, 刘君玲. user trust relationship data.txt. 用户间的信任关系数据.

[3] 薛福亮, 刘君玲. user rating data.txt. 用户评分数据.

[4] 薛福亮, 刘君玲. trust value data.txt. 信任值数据.

[5] 薛福亮, 刘君玲. merge score data.txt. 合并后的评分数据.

[6] 薛福亮, 刘君玲. error analysis data.txt. 误差分析数据.

收稿日期: 2017-05-26
收修改稿日期: 2017-07-16

Improving Collaborative Filtering Recommendation Based on Trust Relationship Among Users

Xue Fuliang　Liu Junling
(Business School, Tianjin University of Finance & Economics, Tianjin 300222, China)

Abstract: [**Objective**] This paper tries to improve user similarity calculation in collaborative filtering recommendation with trust relationship among them. Once there is no similar user for members of the target group, we recommend the most trusted ones as the similar users. [**Methods**] First, we retrieved the trusted users as candidates for the similar users. Second, we combined the trusted and the target users to form a project score set, and evaluated the estimated value of the projects receiving no comment from the target group. Third, we quantified the trust relationship among users to form a regulation factor. Finally, we calculated the adjustment factor and created the similarity cluster of users, and made cross-recommendation among similar users. [**Results**] The collaborative filtering recommendation method based on trust relationship had better performance than traditional ones. [**Limitations**] Only examined the new method with one sample dataset with trusted relationship. More research is needed to test the proposed method with other datasets. [**Conclusions**] The trusted relationship among users contains valuable information, which could be used to calculate user similarity for collaborative filtering recommendation services, and then effectively solves the sparsity and cold start issue.

Keywords: E-commerce Recommendation　User Trust　Collaborative Filtering　Cold Start　Sparsity